Intellectual Property Office
...y and Innovation

Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ
United Kingdom

Telephone +44 (0) 8459 500505
Minicom +44 (0) 8459 222250

Website www.ipo.gov.uk

Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

...he undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) ... the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of ... Comptroller-General, hereby certify that annexed hereto is a true copy of the documents ...originally filed in connection with patent application GB9912635.1 filed on 28 May 1999.

...so certify that the attached copy of the request for grant of a Patent (Form 1/77) bears an ...ndment, effected by this office, following a request by the applicant and agreed to by the ...ptroller-General.

...cordance with the Patents (Companies Re-registration) Rules 1982, if a company named ... ...s certificate and any accompanying documents has re-registered under the Companies ... ...980 with the same name as that with which it was registered immediately before re- ...ration save for the substitution as, or inclusion as, the last part of the name of the words ... ...ic limited company" or their equivalents in Welsh, references to the name of the ... ...any in this certificate and any accompanying documents shall be treated as references to ... ...me with which it is so re-registered.

...rdance with the rules, the words "public limited company" may be replaced by p.l.c., ...C. or PLC.

...stration under the Companies Act does not constitute a new legal entity but merely ... the company to certain additional company law rules.

Signed

Dated 30 April 2007

INVESTOR IN PEOPLE

GB9912635.1

By virtue of a direction given under Section 30 of the Patents Act 1977, the application is proceeding in the name of

GENDAQ LIMITED,
1-3 Burtonhole Lane,
Mill Hill,
LONDON,
NW7 1AD,
Incorporated in the United Kingdom

[ADP No. 07753627001]

Patents Form 1/77

Patents Act 1977
(Rule 16)

The Patent Office

01JUN99 E451088-1 D02246
P01/7700 0.00 - 9912635.1

Request for a grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

28 MAY 1999

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

1. **Your reference** — P006552GB ATM

2. **Patent application number**
(The Patent Office will fill in this part) — **9912635.1**

FILED 3-7-00

3. **Full name, address and postcode of the or of each applicant** *(underline all surnames)*

Medical Research Council
20 Park Crescent
London
W1N 4AL
United Kingdom

Patents ADP number *(if you know it)* — 06007001

If the applicant is a corporate body, give the country/state of its incorporation — United Kingdom

SECTION 8

4. **Title of the invention** — Molecules

5. **Name of your agent** *(if you have one)* — D YOUNG & CO

"Address for service" in the United Kingdom to which all correspondence should be sent
(including the postcode)

21 NEW FETTER LANE
LONDON
EC4A 1DA

Patents ADP number *(if you have one)* — 59006

6. **If you are declaring priority from one or more earlier patent applications, give the country and date of filing of the or each of these earlier applications and (if you know it) the or each application number**

Country	Priority application number *(if you know it)*	Date of filing *(day/month/year)*

7. **If this application is divided or otherwise derived from an earlier UK application, give the number and filing date of the earlier application**

Number of earlier application	Date of filing *(day/month/year)*

8. **Is a statement of inventorship and of right to grant of a patent required in support of this request?** *(Answer 'Yes' if:*
a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d))

Yes

9. **Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document**

Continuation sheets of this form	0
Description	57
Claims(s)	2
Abstract	1
Drawing(s)	2 + 2

10. **If you are also filing any of the following, state how many against each item.**

Priority documents	0
Translations of priority documents	0
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	0
Request for preliminary examination and search *(Patents Form 9/77)*	0
Request for substantive examination *(Patents Form 10/77)*	0
Any other documents *(please specify)*	No

11. **I/We request the grant of a patent on the basis of this application.**

Signature D. Young & Co. **Date** 28.05.99

D YOUNG & CO
Agents for the Applicants

12. **Name and daytime telephone number of the person to contact in the United Kingdom** Dr A Maschio 01703 634816

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 01645 500505

b) Write your answers in capital letters using black ink or you may type them

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you answered 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

Molecules

Field of the Invention

This invention relates to molecules capable of binding DNA in a ligand-dependent manner. Moreover, this invention relates to a method for the identification of said ligand-dependent DNA binding molecules.

Background to the Invention

Gene switches are currently of great interest to those wishing to control timing and/or dosage of gene expression. Various gene switches have been developed in the prior art. Most of these prior art switches are derived from gene regulatory proteins. In these systems, the switching ligand binds to the protein, inducing a protein conformational change that affects DNA binding.

It is often the case that a gene's expression is affected by one or more different protein(s). Diverse proteins may influence expression of the same gene. Said protein(s) may be present in a first cell or cell type, but these protein(s) may be absent from a second cell or cell type. Therefore, a molecule which affects only a single known regulatory protein will not have any effect on the expression of the same gene in a cell where this particular regulatory protein is not expressed, or is otherwise sequestered. Thus, one of the difficulties of the prior art is that a protein-binding switching molecule will have no effect on the expression of a gene if the particular protein to which the switching molecule binds is not present.

Similarly, a gene's expression may be affected by numerous different proteins in different cells or cell types. A molecule which affects only a single known regulatory protein will not have any effect on the expression of the same gene in a cell in which its expression is controlled by a different protein or proteins. Therefore, one of the difficulties in the prior art is that a plurality of switching molecules may be required in order to modulate or switch the expression of a single gene.

Therefore, in order to effect switching of gene expression at a given DNA sequence, independently of the particular activator protein, it is desirable to target the DNA. Further, custom DNA binding proteins would benefit from switches; if these could be designed to interact with DNA, there would be a greater freedom in the design of said proteins.

There are numerous polypeptide modifications which are known to affect their interaction with a broad spectrum of molecules such as nucleic acids, polypeptides (both intra- and inter-molecularly), other macromolecular structures such as membranes, small molecules, ions, or other entities. Clearly, it is a problem that polypeptide modifications may compromise the binding of prior art switching molecules to their polypeptide targets.

The present invention seeks to overcome such difficulties.

Aspects of the present invention are set out in the claims and are described below.

Summary of the Invention

According to the present invention, there is provided a method for isolating a DNA binding molecule which binds to a target DNA molecule in a manner modulatable by a DNA-binding ligand, wherein said DNA-binding ligand and said DNA-binding molecule are different, said method comprising; providing a target DNA molecule; contacting the target DNA molecule with a DNA-binding ligand, to produce a DNA-ligand complex; assessing the ability of candidate DNA-binding molecules to bind the target DNA molecule and the DNA-ligand complex; and isolating those candidate DNA-binding molecules which bind the target DNA molecule and DNA-ligand complex with different binding affinities.

The term 'molecule' is used herein to refer to any atom, ion, molecule, macromolecule (for example polypeptide), or combination of such entities. The term 'ligand' is used interchangeably with the term 'molecule'. Molecules according the invention may be free in solution, or may be partially or fully immobilised. They may be present as discrete entities, or may be complexed with other molecules. Preferably, molecules according to the invention include polypeptides displayed on the surface of bacteriophage particles. More preferably, molecules according to the invention include libraries of polypeptides presented as integral parts of the envelope proteins on the outer surface of bacteriophage particles. Methods for the production of libraries encoding randomised polypeptides are known in the art and may be applied in the present invention. Randomisation may be total, or partial; in the case of partial randomisation, the selected codons preferably encode options for amino acids, and not for stop codons.

Preferably, DNA binding ligands according to the invention may include acridine orange, 9-Amino-6-chloro-2-methoxyacridine, actinomycin D, 7-aminoactinomycin D, dihydroethidium, ethidium-acridine heterodimer, ethidium bromide, propidium iodide, hexidium iodide, Hoechst 33258, Hoechst 33342, hydroxystibamidine, psoralen, Distamycin A, calicheamicin oligosaccharides, triple-helix forming oligos or PNA, pyrole-imidazole polyamides, RNA binding ligands (see below) or any other molecule capable of binding nucleic acid. In a preferred embodiment, derivatives or libraries of said nucleic acid binding ligands may be prepared.

The term 'DNA-binding molecule' includes any molecule which is capable of binding or associating with DNA. This binding or association may be via covalent bonding, via ionic bonding, via hydrogen bonding, via Van-der-Waals bonding, or via any other type of reversible or irreversible association.

DNA binding ligands according to the invention may bind conditionally to nucleic acid. For example, psoralen is a ligand that can bind DNA covalently if illuminated at wavelengths of about 400nm or less. Ligands capable of binding nucleic acids in more

than one manner may be employed in the current invention. Such ligands may bind or associate with the DNA via any one or more mechanism(s) as outlined above.

The term 'complex' is used to describe an association between a DNA and one or more molecules as defined herein.

The term 'candidate DNA-binding molecules' is used to decsribe any one or more molecule(s) as defined above which may or may not be capable of binding DNA. The capability of said molecules to bind DNA may or may not be modulatable by a DNA-binding ligand. These properties may be investigated by the methods of this invention. Preferably, candidate DNA-binding molecules comprise a plurality of, or a library of polypeptides. More preferably, these polypeptides are, or are derived from, DNA-binding proteins such as DNA repair enzymes, polymerases, recombinases, methylases, restriction enzymes, replication factors, histones, or DNA-binding structural proteins such as chromosomal scaffold proteins; even more preferably said polypeptides are derived from transcription factors. 'Derived from' means that the candidate DNA-binding molecules preferably comprise one or more of; transcription factors, fragment(s) of transcription factors, sequences homologous to transcription factors, or polypeptides which have been fully or partially randomised from a starting sequence which is a transcription factor, a fragment of a transcription factor, or homologous to a transcription factor. Most preferably, candidate DNA-binding molecules comprise polypeptides which are at least 40% homologous, more preferably at least 60% homologous, even more preferably at least 75% homologous or even more, for example 85 %, or 90 %, or even more than 95% homologous to one or more transcription factors, using the BLAST algorithm with the parameters as defined below.

In a highly preferred embodiment, these polypeptide candidate DNA-binding molecules are displayed on the surface of bacteriophage particles, and are preferably partially randomised zinc-finger type transcription factors, preferably retaining at least 40% homology (as descibed herein) to zinc-finger type transcription factors.

In some cases, seqeunce homology may be considered in relation to structurally important residues, or those residues which are known or suspected of being evolutionarily conserved. In such instances, residues known to be variable or non-essential for a particular structural conformation may be discounted from the homology calculation. For example, as explained herein, zinc fingers are known to have certain residues which are important for the formation of the three dimensional zinc finger structure. In these cases, homology may be considered over about seven of said important amino acid residues amongst approximately thirty residues which may comprise the whole finger structure.

As used herein, the term homology may refer to structural homology. Structural homology may be estimated by comparing the structural RMS deviation of the main part of the carbon atom backbone of two or more molecules. Preferably, the molecules may be considered structurally homologous if the deviation is 5Å or less, preferably 3Å or less, more preferably 1.5Å or less. Structurally homologous molecules will not necessarily show significant sequence homology.

The term 'target DNA' refers to any DNA for use in the method of the invention. This DNA may be of known sequence, or may be of unknown sequence. This DNA may be prepared artificially in a laboratory, or may be a naturally-occurring DNA. This DNA may be in substantially pure form, or may be in a partially purified form, or may be part of an unpurified or heterogeneous sample. Preferably, the target DNA is a putative promoter. More preferably, the target DNA is in substantially pure form. Even more preferably, the target DNA is of known seqeunce. In a most preferred embodiment, the target DNA is purified DNA of known sequence of a promoter from a gene of interest, preferably from a gene suspected of being associated with a disease state, more preferably from a gene useful in gene therapy.

The term 'modulatable by' is used to indicate that binding of the DNA binding molecule to the DNA can be modulated or affected by the DNA binding ligand. In

other words, the DNA binding ligand can modulate, affect, regulate, adjust, alter, or vary the binding of the DNA binding molecule to the DNA.

Association of the candidate DNA binding molecule with the target DNA may be assessed by any suitable means known to those skilled in the art. For example, the DNA may be immobilised by biotinylation and linking to beads such as streptavidin coated beads (Dynal). In a preferred embodiment wherein the DNA binding molecules are phage displayed polypeptides, binding of said molecules to the DNA may be assessed by eluting those phage which bind, and infecting logarithmic phase *E.coli* TG1 cells. The prescence of infective particles eluted from the DNA indicates that association of the DNA binding molecule(s) with the DNA has occurred. Alternatively, association of the candidate DNA binding molecule(s) with the target DNA may be assessed by Scintillation Proximity Assay (SPA). For example, the target DNA could be biotinylated and immobilised to streptavidin coated SPA beads, and the candidate DNA binding molecules may be radioactively labelled, for example with ^{35}S-Methionine where the molecules are polypeptides. Association of the candidate DNA binding molecules with the target DNA could then be assessed by monitoring the readout of the SPA. Alternatively, the association could be monitored by fluorescent resonance energy transfer (FRET). In this case, the target DNA could be labelled with a donor fluor, and the DNA binding molecule(s) could be labelled with a suitable acceptor flour. Whilst the two entities are seperated, no FRET would be observed, but if association (binding) took place, then there would be a change in the amount of FRET observed, this allowing assessment of the degree of associaiton.

Association of the candidate DNA binding molecule with the target DNA may also be assessed by bandshift assays. Bandshift assays are conducted by measuring the mobility of one or more of the components of the assay, for example the mobility of the DNA, as it is electrophoresed through a suitable gel such as a polyacrylamide acrylamide gel, as is well known to those skilled in the art. In order to assess the association of the candidate DNA binding molecule with the target DNA, the mobility of the DNA could be measured in the prescence and abscence of the candidate DNA

binding molecule. If the mobility of the target DNA is essentially the same in the prescence or abscence of the candidate DNA binding molecule, then it may be inferred that the molecules do not associate, or that the association is weak. If the mobility of the DNA is retarded in the prescence of the candidate DNA binding molecule, then it may be inferred that the candidate molecule is associating with or binding to the DNA.

Association of the candidate DNA binding molecule with the target DNA may also be assessed using filter binding assays. For example, the target DNA molecule may be immobilised on a suitable filter, such as a nitrocellulose filter. The candidate DNA binding molecule may then be labelled, for example radioactively labelled, and contacted with the immobilised target DNA. The binding of or association with the target DNA may be assessed by comparing the amount of labelled candidate DNA binding molecule which associates with the filter only to the amount of labelled candidate DNA binding molecule which associates with the filter-immobilised target DNA. If more labelled candidate DNA binding molecule associates with the immobilised DNA than with the filter only, it may be inferred that the target DNA molecule does indeed associate with the candidate binding molecule.

Binding affinities may be estimated by any suitable means known to those skilled in the art. Binding affinities for the purposes of this invention may be absolute or may be relative. Binding affinities may be determined biochemically, or may simply be estimated by assessing the association of the candidate DNA binding molecule with the target DNA as described above. As used herein, the term binding affinity may refer to a simple estimation of the association of one component of the system with another.

The term 'isolating' in the context of the invention, refers to the act of removing one or more components or molecules from a sample of candidate molecules (as described above) which are used in the methods disclosed herein.

In a second aspect, the invention provides a method for isolating a DNA binding molecule which binds to a target DNA molecule in a manner modulatable by a DNA-binding ligand, wherein said DNA-binding ligand and said DNA-binding molecule are different, and wherein said DNA-binding molecule has a higher affinity for the target DNA in the prescence of ligand than in the abscence of ligand, said method comprising; providing a target DNA molecule; contacting the target DNA molecule with a DNA-binding ligand, to produce a DNA-ligand complex; assessing the ability of candidate DNA-binding moleucles to bind the target DNA molecule and the DNA-ligand complex; and isolating those candidate DNA-binding molecules which bind the DNA-ligand complex with a higher affinity than they bind the target DNA molecule.

In a third aspect, the invention provides a method for isolating a DNA binding molecule which binds to a target DNA molecule in a manner modulatable by a DNA-binding ligand, wherein said DNA-binding ligand and said DNA-binding molecule are different, and wherein said DNA binding molecule binds the target DNA in the abscence of ligand with a higher affinity than it binds the target DNA in the prescence of ligand, said method comprising; providing a target DNA molecule; contacting the target DNA molecule with a DNA-binding ligand, to produce a DNA-ligand complex; assessing the ability of candidate DNA-binding moleucles to bind the target DNA molecule and the DNA-ligand complex, and isolating those candidate DNA-binding molecules which bind the target DNA molecule in the abscence of ligand with a higher affinity than they bind the DNA-ligand complex.

In a preferred aspect of the invention, said candidate molecules are polypeptides.

In a more preferred embodiment, said candidate molecules are polypeptides at least partly derived from transcription factors.

In an even more preferred embodiment, said candidate molecules are derived from zinc finger transcription factors.

Advantageously, the candidate molecules are selected from a phage display library.

In a preferred aspect of the invention, the DNA binding ligand is Distamycin A.

In another aspect, the invention relates to DNA binding molecules obtainable by the methods disclosed herein.

In a further aspect, the invention provides a method of modulating the expression of one or more genes, said method comprising; isolating one or more DNA binding molecule(s) according to any previous claim, and administering said DNA binding molecule(s) to a cell.

According to the invention, candidate DNA binding molecules may comprise, among other things, DNA-binding part(s) of any protein(s), for example zinc finger transcription factors, Zif268, ATF family transcription factors, ATF1, ATF2, bZIP proteins, CHOP, NF-κB, TATA binding protein (TBP), MDM, c-jun, elk, serum response factor (SRF), ternary complex factor (TCF); KRÜPPEL, Odd Skipped, even skipped and other *D.melanogaster* transcription factors; yeast transcription factors such as GCN4, the GAL family of galactose-inducible transcription factors; bacterial transcription factors or repressors such as *lacI*^{q}, or fragments or derivatives thereof. Derivatives would be considered by a person skilled in the art to be functionally and/or structurally related to the molecule(s) from which they are derived, for example through sequence homology of at least 40%.

The DNA-binding molecules according to the invention may be non-randomised polypeptides, for example 'wild-type' or allelic variants of naturally occurring polypeptides, or may be specific mutant(s), or may be wholly or partially randomised polypeptides, preferably structurally related to DNA binding proteins as described herein.

Detailed Description of the Invention

Cys2-His2 zinc finger binding proteins, as is well known in the art, bind to target nucleic acid sequences via α-helical zinc metal atom co-ordinated binding motifs known as zinc fingers. Each zinc finger in a zinc finger nucleic acid binding protein is responsible for determining binding to a nucleic acid triplet, or an overlapping quadruplet, in a nucleic acid binding sequence. Preferably, there are 2 or more zinc fingers, for example 2, 3, 4, 5 or 6 zinc fingers, in each binding protein. Advantageously, there are 3 zinc fingers in each zinc finger binding protein.

All of the DNA-binding residue positions of zinc fingers, as referred to herein, are numbered from the first residue in the α-helix of the finger, ranging from +1 to +9. "-1" refers to the residue in the framework structure immediately preceding the α-helix in a Cys2-His2 zinc finger polypeptide. Residues referred to as "++" are residues present in an adjacent (C-terminal) finger. Where there is no C-terminal adjacent finger, "++" interactions do not operate.

In a first embodiment, the invention provides a method for preparing a DNA binding polypeptide of the Cys2-His2 zinc finger class capable of binding to a target DNA sequence, wherein binding is via a zinc finger DNA binding motif of the polypeptide, and wherein said binding is modulatable by a DNA binding ligand.

The method of the present invention allows the production of what are essentially artificial DNA binding proteins. In these proteins, artificial analogues of amino acids may be used, to impart the proteins with desired properties or for other reasons. Thus, the term "amino acid", particularly in the context where "any amino acid" is referred to, means any sort of natural or artificial amino acid or amino acid analogue that may be employed in protein construction according to methods known in the art. Moreover, any specific amino acid referred to herein may be replaced by a functional analogue thereof, particularly an artificial functional analogue. The nomenclature used herein therefore specifically comprises within its scope functional analogues or mimetics of the defined amino acids.

The α-helix of a zinc finger binding protein aligns antiparallel to the nucleic acid strand, such that the primary nucleic acid sequence is arranged 3' to 5' in order to correspond with the N terminal to C-terminal sequence of the zinc finger. Since nucleic acid sequences are conventionally written 5' to 3', and amino acid sequences N-terminus to C-terminus, the result is that when a nucleic acid sequence and a zinc finger protein are aligned according to convention, the primary interaction of the zinc finger is with the - strand of the nucleic acid, since it is this strand which is aligned 3' to 5'. These conventions are followed in the nomenclature used herein. It should be noted, however, that in nature certain fingers, such as finger 4 of the protein GLI, bind to the + strand of nucleic acid: see Suzuki *et al.*, (1994) NAR 22:3397-3405 and Pavletich and Pabo, (1993) Science 261:1701-1707. The incorporation of such fingers into DNA binding molecules according to the invention is envisaged.

The present invention may be integrated with the rules set forth for zinc finger polypeptide design in our copending European or PCT patent applications having publication numbers; WO 98/53057, WO 98/53060, WO 98/53058, WO 98/53059, describe improved techniques for designing zinc finger polypeptides capable of binding desired nucleic acid sequences. In combination with selection procedures, such as phage display, set forth for example in WO 96/06166, these techniques enable the production of zinc finger polypeptides capable of recognising practically any desired sequence.

In a preferred aspect, therefore, the invention provides a method for preparing a DNA binding polypeptide of the Cys2-His2 zinc finger class capable of binding to a target DNA sequence, wherein said binding is modulatable by a DNA binding ligand, and wherein binding to each base of the triplet by an α-helical zinc finger DNA binding motif in the polypeptide is determined as follows:

a) if the 5' base in the triplet is G, then position +6 in the α-helix is Arg and/or position ++2 is Asp;

b) if the 5' base in the triplet is A, then position +6 in the α-helix is Gln or Glu and ++2 is not Asp;

c) if the 5' base in the triplet is T, then position +6 in the α-helix is Ser or Thr and position ++2 is Asp; or position +6 is a hydrophobic amino acid other than Ala;

d) if the 5' base in the triplet is C, then position +6 in the α-helix may be any amino acid, provided that position ++2 in the α-helix is not Asp;

e) if the central base in the triplet is G, then position +3 in the α-helix is His;

f) if the central base in the triplet is A, then position +3 in the α-helix is Asn;

g) if the central base in the triplet is T, then position +3 in the α-helix is Ala, Ser, Ile, Leu, Thr or Val; provided that if it is Ala, then one of the residues at -1 or +6 is a small residue;

h) if the central base in the triplet is 5-meC, then position +3 in the α-helix is Ala, Ser, Ile, Leu, Thr or Val; provided that if it is Ala, then one of the residues at -1 or +6 is a small residue;

i) if the 3' base in the triplet is G, then position -1 in the α-helix is Arg;

j) if the 3' base in the triplet is A, then position -1 in the α-helix is Gln and position +2 is Ala;

k) if the 3' base in the triplet is T, then position -1 in the α-helix is Asn; or position -1 is Gln and position +2 is Ser;

l) if the 3' base in the triplet is C, then position -1 in the α-helix is Asp and Position +1 is Arg; where the central residue of a target triplet is C, the use of Asp at position +3 of a zinc finger polypeptide allows preferential binding to C over 5-meC.

The foregoing represents a set of rules which permits the design of a zinc finger binding protein specific for any given target DNA sequence.

A zinc finger binding motif is a structure well known to those in the art and defined in, for example, Miller *et al.*, (1985) EMBO J. 4:1609-1614; Berg (1988) PNAS (USA) 85:99-102; Lee *et al.*, (1989) Science 245:635-637; see International patent

applications WO 96/06166 and WO 96/32475, corresponding to USSN 08/422,107, incorporated herein by reference.

In general, a preferred zinc finger framework has the structure:

(A) X_{0-2} **C** X_{1-5} **C** X_{9-14} **H** X_{3-6} **H**/**C**

where X is any amino acid, and the numbers in subscript indicate the possible numbers of residues represented by X.

In a preferred aspect of the present invention, zinc finger nucleic acid binding motifs may be represented as motifs having the following primary structure:

(B) X^a **C** X_{2-4} **C** X_{2-3} F X^c X X X X L X X **H** X X X^b **H** - linker

-1 1 2 3 4 5 6 7 8 9

wherein X (including X^a, X^b and X^c) is any amino acid. X_{2-4} and X_{2-3} refer to the presence of 2 or 4, or 2 or 3, amino acids, respectively. The Cys and His residues, which together co-ordinate the zinc metal atom, are marked in bold text and are usually invariant, as is the Leu residue at position +4 in the α-helix.

Modifications to this representation may occur or be effected without necessarily abolishing zinc finger function, by insertion, mutation or deletion of amino acids. For example it is known that the second His residue may be replaced by Cys (Krizek *et al.*, (1991) J. Am. Chem. Soc. 113:4518-4523) and that Leu at +4 can in some circumstances be replaced with Arg. The Phe residue before X_c may be replaced by any aromatic other than Trp. Moreover, experiments have shown that departure from the preferred structure and residue assignments for the zinc finger are tolerated and may even prove beneficial in binding to certain nucleic acid sequences. Even taking this into account, however, the general structure involving an α-helix co-ordinated by a zinc atom which contacts four Cys or His residues, does not alter. As used herein,

structures (A) and (B) above are taken as an exemplary structure representing all zinc finger structures of the Cys2-His2 type.

Preferably, X^a is $^F/_Y$-X or P-$^F/_Y$-X. In this context, X is any amino acid. Preferably, in this context X is E, K, T or S. Less preferred but also envisaged are Q, V, A and P. The remaining amino acids remain possible.

Preferably, X_{2-4} consists of two amino acids rather than four. The first of these amino acids may be any amino acid, but S, E, K, T, P and R are preferred. Advantageously, it is P or R. The second of these amino acids is preferably E, although any amino acid may be used.

Preferably, X^b is T or I.

Preferably, X^c is S or T.

Preferably, X_{2-3} is G-K-A, G-K-C, G-K-S or G-K-G. However, departures from the preferred residues are possible, for example in the form of M-R-N or M-R.

Preferably, the linker is T-G-E-K or T-G-E-K-P.

As set out above, the major binding interactions occur with amino acids -1, +3 and +6. Amino acids +4 and +7 are largely invariant. The remaining amino acids may be essentially any amino acids. Preferably, position +9 is occupied by Arg or Lys. Advantageously, positions +1, +5 and +8 are not hydrophobic amino acids, that is to say are not Phe, Trp or Tyr. Preferably, position ++2 is any amino acid, and preferably serine, save where its nature is dictated by its role as a ++2 amino acid for an N-terminal zinc finger in the same nucleic acid binding molecule.

In a most preferred aspect, therefore, bringing together the above, the invention allows the definition of every residue in a zinc finger DNA binding motif which will bind specifically to a given target DNA triplet.

The code provided by the present invention is not entirely rigid; certain choices are provided. For example, positions +1, +5 and +8 may have any amino acid allocation, whilst other positions may have certain options: for example, the present rules provide that, for binding to a central T residue, any one of Ala, Ser or Val may be used at +3. In its broadest sense, therefore, the present invention provides a very large number of proteins which are capable of binding to every defined target DNA triplet.

Preferably, however, the number of possibilities may be significantly reduced. For example, the non-critical residues +1, +5 and +8 may be occupied by the residues Lys, Thr and Gln respectively as a default option. In the case of the other choices, for example, the first-given option may be employed as a default. Thus, the code according to the present invention allows the design of a single, defined polypeptide (a "default" polypeptide) which will bind to its target triplet.

In a further aspect of the present invention, there is provided a method for preparing a DNA binding protein of the Cys2-His2 zinc finger class capable of binding to a target DNA sequence in a manner modulatable by a DNA binding ligand, comprising the steps of:

a) selecting a model zinc finger domain from the group consisting of naturally occurring zinc fingers and consensus zinc fingers; and

b) mutating at least one of positions -1, +3, +6 (and ++2) of the finger as required by a method according to the present invention.

In general, naturally occurring zinc fingers may be selected from those fingers for which the DNA binding specificity is known. For example, these may be the fingers

for which a crystal structure has been resolved: namely Zif 268 (Elrod-Erickson *et al.*, (1996) Structure 4:1171-1180), GLI (Pavletich and Pabo, (1993) Science 261:1701-1707), Tramtrack (Fairall *et al.*, (1993) Nature 366:483-487) and YY1 (Houbaviy *et al.*, (1996) PNAS (USA) 93:13577-13582).

The naturally occurring zinc finger 2 in Zif 268 makes an excellent starting point from which to engineer a zinc finger and is preferred.

Consensus zinc finger structures may be prepared by comparing the sequences of known zinc fingers, irrespective of whether their binding domain is known. Preferably, the consensus structure is selected from the group consisting of the consensus structure P Y K C P E C G K S F S Q K S D L V K H Q R T H T G, and the consensus structure P Y K C S E C G K A F S Q K S N L T R H Q R I H T G E K P.

The consensuses are derived from the consensus provided by Krizek *et al.*, (1991) J. Am. Chem. Soc. 113:4518-4523 and from Jacobs, (1993) PhD thesis, University of Cambridge, UK. In both cases, the linker sequences described above for joining two zinc finger motifs together, namely TGEK or TGEKP can be formed on the ends of the consensus. Thus, a P may be removed where necessary, or, in the case of the consensus terminating T G, E K (P) can be added.

When the nucleic acid specificity of the model finger selected is known, the mutation of the finger in order to modify its specificity to bind to the target DNA may be directed to residues known to affect binding to bases at which the natural and desired targets differ. Otherwise, mutation of the model fingers should be concentrated upon residues -1, +3, +6 and ++2 as provided for in the foregoing rules.

In order to produce a binding protein having improved binding, moreover, the rules provided by the present invention may be supplemented by physical or virtual modelling of the protein/DNA interface in order to assist in residue selection.

In a second embodiment, the invention provides a method for producing a zinc finger polypeptide capable of binding to a target DNA sequence, wherein said binding is modulatable by a DNA binding ligand, comprising:

a) providing a nucleic acid library encoding a repertoire of zinc finger polypeptides, the nucleic acid members of the library being at least partially randomised at one or more of the positions encoding residues -1, 2, 3 and 6 of the α-helix of the zinc finger polypeptides;

b) displaying the library in a selection system and screening it against a target DNA sequence;

c) isolating the nucleic acid members of the library encoding zinc finger polypeptides capable of binding to the target sequence in the prescence/abscence of DNA binding ligand;

d) selecting those members of the library isolated in (c) which bind the target nucleic acid sequence with different affinities in the prescence and abscence of the DNA binding ligand.

Methods for the production of libraries encoding randomised polypeptides are known in the art and may be applied in the present invention. Randomisation may be total, or partial; in the case of partial randomisation, the selected codons preferably encode options for amino acids as set forth in the rules of the first embodiment of the present invention. Thus, the first and second embodiments may advantageously be combined.

Zinc finger polypeptides may be designed which specifically bind to nucleic acids incorporating the base U, in preference to the equivalent base T.

In a further preferred aspect, the invention comprises a method for producing a zinc finger polypeptide capable of binding to a target DNA sequence, wherein said binding is modulatable by a DNA binding ligand, comprising:

a) providing a nucleic acid library encoding a repertoire of zinc finger polypeptides each possessing more than one zinc fingers, the nucleic acid members of the library being at least partially randomised at one or more of the positions encoding residues -1, 2, 3 and 6 of the α-helix in a first zinc finger and at one or more of the positions encoding residues -1, 2, 3 and 6 of the α-helix in a further zinc finger of the zinc finger polypeptides;

b) displaying the library in a selection system and screening it against a target DNA sequence;

c) assessing the affinity of the DNA binding molecules for the target DNA in the prescence and abscence of the DNA-binding ligand, and

d) isolating the nucleic acid members of the library encoding zinc finger polypeptides capable of binding to the target sequence with different affinities in the prescence and abscence of DNA binding ligand.

In this aspect, the invention encompasses library technology described in our copending International patent application WO 98/53057, incorporated herein by reference in its entirety. WO 98/53057 describes the production of zinc finger polypeptide libraries in which each individual zinc finger polypeptide comprises more than one, for example two or three, zinc fingers; and wherein within each polypeptide partial randomisation occurs in at least two zinc fingers.

This allows for the selection of the "overlap" specificity, wherein, within each triplet, the choice of residue for binding to the third nucleotide (read 3' to 5' on the + strand) is influenced by the residue present at position +2 on the subsequent zinc finger, which

displays cross-strand specificity in binding. The selection of zinc finger polypeptides incorporating cross-strand specificity of adjacent zinc fingers enables the selection of nucleic acid binding proteins more quickly, and/or with a higher degree of specificity than is otherwise possible.

The present invention relates to a method for engineering a novel class of gene switches in which a DNA binding ligand affects or modulates the interaction of a DNA binding molecule (for example phage displayed polypeptide), with its target DNA. In a preferred aspect, the present invention relates to the selection of DNA-binding polypeptides which recognise a particular DNA sequence or structure. Preferably, said method may include selection of phage displayed polypeptides that bind a DNA target in the presence or abscence of one or more DNA binding ligands. Of the phage displayed polypeptides which are selected under these conditions, some may bind the DNA with higher affinity in the prescence of ligand, whereas others may bind the DNA with higher affinity in the abscence of ligand.

In order to remove DNA binding molecules (for example phage displayed polypeptides) which bind DNA in a ligand-independent manner from a library, a pre-selection step may optionally be performed in the absence of ligand prior to each round of selection. This step removes from the library those clones which do not require ligand for DNA binding. Optionally, candidate molecules selected in this manner may be screened by ELISA for binding to the DNA target in the presence or absence of the ligand(s).

In order to remove DNA binding molecules (for example phage displayed polypeptides) which bind DNA in a ligand-dependent manner from a library, a pre-selection step may optionally be performed in the presence of ligand prior to each round of selection. This step removes from the library those clones which require ligand for DNA binding. Optionally, candidate molecules selected in this manner may be screened by ELISA for binding to the DNA target in the presence or absence of the ligand(s).

Preferably, the association of the DNA binding molecule with a target nucleic acid can be affected or modulated (switched) by a DNA binding ligand.

Randomisation may involve alteration of zinc finger polypeptides, said alteration being accomplished at the DNA or protein level. Mutagenesis and screening of zinc finger polypeptides may be achieved by any suitable means. Preferably, the mutagenesis is performed at the nucleic acid level, for example by synthesising novel genes encoding mutant polypeptides and expressing these to obtain a variety of different proteins. Alternatively, existing genes can themselves be mutated, such as by site-directed or random mutagenesis, in order to obtain the desired mutant genes.

Mutations may be performed by any method known to those of skill in the art. Preferred, however, is site-directed mutagenesis of a nucleic acid sequence encoding the protein of interest. A number of methods for site-directed mutagenesis are known in the art, from methods employing single-stranded phage such as M13 to PCR-based techniques (see "PCR Protocols: A guide to methods and applications", M.A. Innis, D.H. Gelfand, J.J. Sninsky, T.J. White (eds.). Academic Press, New York, 1990). Preferably, the commercially available Altered Site II Mutagenesis System (Promega) may be employed, according to the manufacturer's instructions.

Randomisation of the zinc finger binding motifs produced according to the invention is preferably directed to those amino acid residues where the code provided herein gives a choice of residues (see above). For example, positions +1, +5 and +8 are advantageously randomised, whilst preferably avoiding hydrophobic amino acids; positions involved in binding to the nucleic acid, notably -1, +2, +3 and +6, may be randomised also, preferably within the choices provided by the rules of the present invention.

Screening of the proteins produced by mutant genes is preferably performed by expressing the genes and assaying the binding ability of the protein product. A simple

and advantageously rapid method by which this may be accomplished is by phage display, in which the mutant polypeptides are expressed as fusion proteins with the coat proteins of filamentous bacteriophage, such as the minor coat protein pII of bacteriophage m13 or gene III of bacteriophage Fd, and displayed on the capsid of bacteriophage transformed with the mutant genes. The target nucleic acid sequence is used as a probe to bind directly to the protein on the phage surface and select the phage possessing advantageous mutants, by affinity purification. The phage are then amplified by passage through a bacterial host, and subjected to further rounds of selection and amplification in order to enrich the mutant pool for the desired phage and eventually isolate the preferred clone(s). Detailed methodology for phage display is known in the art and set forth, for example, in US Patent 5,223,409; Choo and Klug, (1995) Current Opinions in Biotechnology 6:431-436; Smith, (1985) Science 228:1315-1317; and McCafferty *et al.*, (1990) Nature 348:552-554; all incorporated herein by reference. Vector systems and kits for phage display are available commercially, for example from Pharmacia.

Specific peptide ligands such as zinc finger polypeptides may moreover be selected for binding to targets by affinity selection using large libraries of peptides linked to the C terminus of the lac repressor LacI (Cull *et al.*, (1992) Proc Natl Acad Sci U S A, 89, 1865-9). When expressed in *E. coli* the repressor protein physically links the ligand to the encoding plasmid by binding to a lac operator sequence on the plasmid.

An entirely *in vitro* polysome display system has also been reported (Mattheakis *et al.*, (1994) Proc Natl Acad Sci U S A, 91, 9022-6) in which nascent peptides are physically attached via the ribosome to the RNA which encodes them. Furthermore, polypeptides may be partitioned in physical compartments for example wells of an *in vitro* dish, or subcellular compartments, or in small fluid particles or droplets such as emulsions; further teachings on this topic may be found in Griff *et al.*, (see WO 99/02671).

A library of the invention may be randomised at those positions for which choices are given in the rules of the first embodiment of the present invention. The rules set forth

above allow the person of ordinary skill in the art to make informed choices concerning the desired codon usage at the given positions.

Zinc finger binding motifs designed according to the invention may be combined into nucleic acid binding polypeptide molecules having a multiplicity of zinc fingers. Preferably, the proteins have at least two zinc fingers. In nature, zinc finger binding proteins commonly have at least three zinc fingers, although two-zinc finger proteins such as Tramtrack are known. The presence of at least three zinc fingers is preferred. Nucleic acid binding proteins may be constructed by joining the required fingers end to end, N-terminus to C-terminus. Preferably, this is effected by joining together the relevant nucleic acid sequences which encode the zinc fingers to produce a composite nucleic acid coding sequence encoding the entire binding protein. The invention therefore provides a method for producing a DNA binding protein as defined above, wherein the DNA binding protein is constructed by recombinant DNA technology, the method comprising the steps of:

a) preparing a nucleic acid coding sequence encoding two or more zinc finger binding motifs as defined above, placed N-terminus to C-terminus;

b) inserting the nucleic acid sequence into a suitable expression vector; and

c) expressing the nucleic acid sequence in a host organism in order to obtain the DNA binding protein.

A "leader" peptide may be added to the N-terminal finger. Preferably, the leader peptide is MAEEKP.

The nucleic acid encoding the DNA binding protein according to the invention can be incorporated into vectors for further manipulation. As used herein, vector (or plasmid) refers to discrete elements that are used to introduce heterologous nucleic acid into cells for either expression or replication thereof. Selection and use of such vehicles are well within the skill of the person of ordinary skill in the art. Many vectors are available, and selection of appropriate vector will depend on the intended use of the

vector, i.e. whether it is to be used for DNA amplification or for nucleic acid expression, the size of the DNA to be inserted into the vector, and the host cell to be transformed with the vector. Each vector contains various components depending on its function (amplification of DNA or expression of DNA) and the host cell for which it is compatible. The vector components generally include, but are not limited to, one or more of the following: an origin of replication, one or more marker genes, an enhancer element, a promoter, a transcription termination sequence and a signal sequence.

Both expression and cloning vectors generally contain nucleic acid sequence that enable the vector to replicate in one or more selected host cells. Typically in cloning vectors, this sequence is one that enables the vector to replicate independently of the host chromosomal DNA, and includes origins of replication or autonomously replicating sequences. Such sequences are well known for a variety of bacteria, yeast and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2μ plasmid origin is suitable for yeast, and various viral origins (e.g. SV 40, polyoma, adenovirus) are useful for cloning vectors in mammalian cells. Generally, the origin of replication component is not needed for mammalian expression vectors unless these are used in mammalian cells competent for high level DNA replication, such as COS cells.

Most expression vectors are shuttle vectors, i.e. they are capable of replication in at least one class of organisms but can be transfected into another class of organisms for expression. For example, a vector is cloned in *E.coli* and then the same vector is transfected into yeast or mammalian cells even though it is not capable of replicating independently of the host cell chromosome. DNA may also be replicated by insertion into the host genome. However, the recovery of genomic DNA encoding the DNA binding protein is more complex than that of episomally replicated vector because restriction enzyme digestion is required to excise DNA binding protein DNA. DNA can be amplified by PCR and be directly transfected into the host cells without any replication component.

Advantageously, an expression and cloning vector may contain a selection gene also referred to as selectable marker. This gene encodes a protein necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that confer resistance to antibiotics and other toxins, e.g. ampicillin, neomycin, methotrexate or tetracycline, complement auxotrophic deficiencies, or supply critical nutrients not available from complex media.

Selectable markers which may be used in fungal cells, for example yeast cells, include wild-type genes which complement auxotrophic defects in for example the Uracil (eg. URA3 gene), Lysine (eg. LYS2 gene), Adenine (eg. ADE2 gene), Methionine (eg. MET3 gene), Histidine (eg. HIS3 gene), Tryptophan (eg. TRP1 gene), Leucine (eg. LEU2 gene) or other metabolic pathways. In addition, counter-selection methods are well known in the art. These enable genes to be selected against by the action of a chemical precursor which is harmless unless converted to a toxic product by the action of one or more gene(s). Examples of these include; 5-fluoro-orotic acid, which is converted to a toxic compound by the action of the URA3 gene product; α-amino-adipic acid, which is converted to a toxic compound by the LYS2 gene product; allyl alcohol, which is converted to a toxic compound by alcohol dehydrogenase activity as encoded by the ADH genes, or any other suitable selective regime known to those skilled in the art. Other selective markers are based on the expression of a gene in a fungus such as yeast which overcomes the metabolic arrest induced by, or toxicity of, a chemical entity which may be added to the growth medium or otherwise presented to the cells. Examples of these may include the KAN gene(s) which confer resistance to antibiotics such as G-148, the HIS3 gene which confers resistance to 3-amino-triazole, or the ADH2 gene which can confer resistance to heavy metal ions such as cadmium, or any other suitable genes which confer resistance to toxic or growth arresting regimes.

Since the replication of vectors is conveniently done in *E.coli*, an *E.coli* genetic marker and an *E.coli* origin of replication are advantageously included. These can be obtained from *E.coli* plasmids, such as pBR322, Bluescript© vector or a pUC plasmid, e.g. pUC18 or pUC19, which contain both *E.coli* replication origin and *E.coli* genetic marker conferring resistance to antibiotics, such as ampicillin.

Suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up DNA binding protein nucleic acid, such as dihydrofolate reductase (DHFR, methotrexate resistance), thymidine kinase, or genes conferring resistance to G418 or hygromycin. The mammalian cell transformants are placed under selection pressure which only those transformants which have taken up and are expressing the marker are uniquely adapted to survive. In the case of a DHFR or glutamine synthase (GS) marker, selection pressure can be imposed by culturing the transformants under conditions in which the pressure is progressively increased, thereby leading to amplification (at its chromosomal integration site) of both the selection gene and the linked DNA that encodes the DNA binding protein. Amplification is the process by which genes in greater demand for the production of a protein critical for growth, together with closely associated genes which may encode a desired protein, are reiterated in tandem within the chromosomes of recombinant cells. Increased quantities of desired protein are usually synthesised from thus amplified DNA.

Expression and cloning vectors usually contain a promoter that is recognised by the host organism and is operably linked to nucleic acid encoding DNA binding protein. Such a promoter may be inducible or constitutive. The promoters are operably linked to DNA encoding the DNA binding protein by removing the promoter from the source DNA by restriction enzyme digestion and inserting the isolated promoter sequence into the vector. Both the native DNA binding protein promoter sequence and many heterologous promoters may be used to direct amplification and/or expression of DNA binding protein encoding DNA.

Promoters suitable for use with prokaryotic hosts include, for example, the β-lactamase and lactose promoter systems, alkaline phosphatase, the tryptophan (trp) promoter system and hybrid promoters such as the tac promoter. Their nucleotide sequences have been published, thereby enabling the skilled worker operably to ligate them to DNA encoding DNA binding protein, using linkers or adapters to supply any required restriction sites. Promoters for use in bacterial systems will also generally contain a Shine-Delgarno sequence operably linked to the DNA encoding the DNA binding protein.

Preferred expression vectors are bacterial expression vectors which comprise a promoter of a bacteriophage such as phagex or T7 which is capable of functioning in the bacteria. In one of the most widely used expression systems, the nucleic acid encoding the fusion protein may be transcribed from the vector by T7 RNA polymerase (Studier et al, Methods in Enzymol. 185; 60-89, 1990). In the *E.coli* BL21(DE3) host strain, used in conjunction with pET vectors, the T7 RNA polymerase is produced from the λ-lysogen DE3 in the host bacterium, and its expression is under the control of the IPTG inducible lac UV5 promoter. This system has been employed successfully for over-production of many proteins. Alternatively the polymerase gene may be introduced on a lambda phage by infection with an int- phage such as the CE6 phage which is commercially available (Novagen, Madison, USA). other vectors include vectors containing the lambda PL promoter such as PLEX (Invitrogen, NL), vectors containing the trc promoters such as pTrcHisXpressTm (Invitrogen) or pTrc99 (Pharmacia Biotech, SE) or vectors containing the tac promoter such as pKK223-3 (Pharmacia Biotech) or PMAL (New England Biolabs, MA, USA).

Moreover, the DNA binding protein gene according to the invention preferably includes a secretion sequence in order to facilitate secretion of the polypeptide from bacterial hosts, such that it will be produced as a soluble native peptide rather than in an inclusion body. The peptide may be recovered from the bacterial periplasmic space, or the culture medium, as appropriate.

Suitable promoting sequences for use with yeast hosts may be regulated or constitutive and are preferably derived from a highly expressed yeast gene, especially a Saccharomyces cerevisiae gene. Thus, the promoter of the TRP1 gene, the ADHI or ADHII gene, the acid phosphatase (PH05) gene, a promoter of the yeast mating pheromone genes coding for the a- or α-factor or a promoter derived from a gene encoding a glycolytic enzyme such as the promoter of the enolase, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), 3-phospho glycerate kinase (PGK), hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triose phosphate isomerase, phosphoglucose isomerase or glucokinase genes, or a promoter from the TATA binding protein (TBP) gene can be used. Furthermore, it is possible to use hybrid promoters comprising upstream activation sequences (UAS) of one yeast gene and downstream promoter elements including a functional TATA box of another yeast gene, for example a hybrid promoter including the UAS(s) of the yeast PH05 gene and downstream promoter elements including a functional TATA box of the yeast GAP gene (PH05-GAP hybrid promoter). A suitable constitutive PHO5 promoter is e.g. a shortened acid phosphatase PH05 promoter devoid of the upstream regulatory elements (UAS) such as the PH05 (-173) promoter element starting at nucleotide -173 and ending at nucleotide -9 of the PH05 gene.

DNA binding protein gene transcription from vectors in mammalian hosts may be controlled by promoters derived from the genomes of viruses such as polyoma virus, adenovirus, fowlpox virus, bovine papilloma virus, avian sarcoma virus, cytomegalovirus (CMV), a retrovirus and Simian Virus 40 (SV40), from heterologous mammalian promoters such as the actin promoter or a very strong promoter, e.g. a ribosomal protein promoter, and from the promoter normally associated with DNA binding protein sequence, provided such promoters are compatible with the host cell systems.

Transcription of a DNA encoding DNA binding protein by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are relatively

orientation and position independent. Many enhancer sequences are known from mammalian genes (e.g. elastase and globin). However, typically one will employ an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270) and the CMV early promoter enhancer. The enhancer may be spliced into the vector at a position 5' or 3' to DNA binding protein DNA, but is preferably located at a site 5' from the promoter.

Advantageously, a eukaryotic expression vector encoding a DNA binding protein according to the invention may comprise a locus control region (LCR). LCRs are capable of directing high-level integration site independent expression of transgenes integrated into host cell chromatin, which is of importance especially where the DNA binding protein gene is to be expressed in the context of a permanently-transfected eukaryotic cell line in which chromosomal integration of the vector has occurred, or in transgenic animals.

Eukaryotic vectors may also contain sequences necessary for the termination of transcription and for stabilising the mRNA. Such sequences are commonly available from the 5' and 3' untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding DNA binding protein.

An expression vector includes any vector capable of expressing DNA binding protein nucleic acids that are operatively linked with regulatory sequences, such as promoter regions, that are capable of expression of such DNAs. Thus, an expression vector refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector, that upon introduction into an appropriate host cell, results in expression of the cloned DNA. Appropriate expression vectors are well known to those with ordinary skill in the art and include those that are replicable in eukaryotic and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome. For example, DNAs encoding DNA binding protein may be inserted into a vector suitable for expression of cDNAs in mammalian

cells, e.g. a CMV enhancer-based vector such as pEVRF (Matthias, et al., (1989) NAR 17, 6418).

Particularly useful for practising the present invention are expression vectors that provide for the transient expression of DNA encoding DNA binding protein in mammalian cells. Transient expression usually involves the use of an expression vector that is able to replicate efficiently in a host cell, such that the host cell accumulates many copies of the expression vector, and, in turn, synthesises high levels of DNA binding protein. For the purposes of the present invention, transient expression systems are useful e.g. for identifying DNA binding protein mutants, to identify potential phosphorylation sites, or to characterise functional domains of the protein.

Construction of vectors according to the invention employs conventional ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required. If desired, analysis to confirm correct sequences in the constructed plasmids is performed in a known fashion. Suitable methods for constructing expression vectors, preparing in vitro transcripts, introducing DNA into host cells, and performing analyses for assessing DNA binding protein expression and function are known to those skilled in the art. Gene presence, amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA, dot blotting (DNA or RNA analysis), or in situ hybridisation, using an appropriately labelled probe which may be based on a sequence provided herein. Those skilled in the art will readily envisage how these methods may be modified, if desired.

In accordance with another embodiment of the present invention, there are provided cells containing the above-described nucleic acids. Such host cells such as prokaryote, yeast and higher eukaryote cells may be used for replicating DNA and producing the DNA binding protein. Suitable prokaryotes include eubacteria, such as Gram-negative

or Gram-positive organisms, such as *E.coli*, e.g. *E.coli* K-12 strains, DH5α and HB101, or Bacilli. Further hosts suitable for the DNA binding protein encoding vectors include eukaryotic microbes such as filamentous fungi or yeast, e.g. Saccharomyces cerevisiae. Higher eukaryotic cells include insect and vertebrate cells, particularly mammalian cells including human cells, or nucleated cells from other multicellular organisms. In recent years propagation of vertebrate cells in culture (tissue culture) has become a routine procedure. Examples of useful mammalian host cell lines are epithelial or fibroblastic cell lines such as Chinese hamster ovary (CHO) cells, NIH 3T3 cells, HeLa cells or 293T cells. The host cells referred to in this disclosure comprise cells in *in vitro* culture as well as cells that are within a host animal.

DNA may be stably incorporated into cells or may be transiently expressed using methods known in the art. Stably transfected mammalian cells may be prepared by transfecting cells with an expression vector having a selectable marker gene, and growing the transfected cells under conditions selective for cells expressing the marker gene. To prepare transient transfectants, mammalian cells are transfected with a reporter gene to monitor transfection efficiency.

To produce such stably or transiently transfected cells, the cells should be transfected with a sufficient amount of the DNA binding protein-encoding nucleic acid to form the DNA binding protein. The precise amounts of DNA encoding the DNA binding protein may be empirically determined and optimised for a particular cell and assay.

Host cells are transfected or, preferably, transformed with the above-mentioned expression or cloning vectors of this invention and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. Heterologous DNA may be introduced into host cells by any method known in the art, such as transfection with a vector encoding a heterologous DNA by the calcium phosphate coprecipitation technique or by electroporation. Numerous methods of transfection are known to the

skilled worker in the field. Successful transfection is generally recognised when any indication of the operation of this vector occurs in the host cell. Transformation is achieved using standard techniques appropriate to the particular host cells used.

Incorporation of cloned DNA into a suitable expression vector, transfection of eukaryotic cells with a plasmid vector or a combination of plasmid vectors, each encoding one or more distinct genes or with linear DNA, and selection of transfected cells are well known in the art (see, e.g. Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press).

Transfected or transformed cells are cultured using media and culturing methods known in the art, preferably under conditions whereby the DNA binding protein encoded by the DNA is expressed. The composition of suitable media is known to those in the art, so that they can be readily prepared. Suitable culturing media are also commercially available.

DNA binding proteins according to the invention may be employed in a wide variety of applications, including diagnostics and as research tools. Advantageously, they may be employed as diagnostic tools for identifying the presence of particular nucleic acid molecules in a complex mixture. DNA binding molecules according to the invention can preferably differentiate between different target DNA molecules, and their binding affinities for the DNA target sequences are preferably modulated by DNA binding ligand(s). DNA binding molecules according to the invention are useful in switching or modulating gene expression, especially in gene therapy applications.

For example, zinc fingers may be fused to nucleic acid cleavage moieties, such as the catalytic domain of a restriction enzyme, to produce a restriction enzyme capable of cleaving only target DNA of a specific sequence (see Kim, *et al.*, (1996) Proc. Natl. Acad. Sci. USA 93:1156-1160). Using such approaches, different zinc finger domains can be used to create restriction enzymes with any desired recognition

nucleotide sequence, but which cleave DNA conditionally dependent on the prescence or abscence of a particular DNA binding ligand, for instance Distamycin A.

Targeted zinc fingers according to the invention may moreover be employed in the regulation of gene transcription, for example by specific cleavage of nucleic acid sequences using a fusion polypeptide comprising a zinc finger targeting domain and a DNA cleavage domain, or by fusion of an activating domain (such as HSV VP16) to a zinc finger, to activate transcription from a gene which possesses the zinc finger binding sequence in its upstream sequences. Preferably, activation only occurs in the prescence of the DNA binding ligand, since the zinc fingers will not bind their target nucleic acid sequences in the abscence of the ligand. Alternatively, activation only occurs in the abscence of the DNA binding ligand, since the zinc fingers may not bind their target nucleic acid sequences in the prescence of the ligand. Zinc fingers capable of differentiating between U and T may be used to preferentially target RNA or DNA, as required. Where RNA-targeting polypeptides are intended, these are included in the term "DNA-binding molecule".

In a preferred embodiment, the zinc finger polypeptides of the invention may be employed to detect the presence of a particular target nucleic acid sequence in a sample.

Accordingly, the invention provides a method for determining the presence of a target nucleic acid molecule, comprising the steps of:

a) preparing a DNA binding protein by the method set forth above which is specific for the target nucleic acid molecule;

b) exposing a test system which may comprise the target nucleic acid molecule to the DNA binding protein under conditions which promote binding, and removing any DNA binding protein which remains unbound;

c) detecting the presence of the DNA binding protein in the test system.

In a preferred embodiment, the DNA binding molecules of the invention can be incorporated into an ELISA assay. For example, phage displaying the molecules of the invention can be used to detect the presence of the target DNA, and visualised using enzyme-linked anti-phage antibodies.

Further improvements to the use of zinc finger phage for diagnosis can be made, for example, by co-expressing a marker protein fused to the minor coat protein (gVIII) of bacteriophage. Since detection with an anti-phage antibody would then be obsolete, the time and cost of each diagnosis would be further reduced. Depending on the requirements, suitable markers for display might include the fluorescent proteins (A. B. Cubitt, *et al.*, (1995) *Trends Biochem Sci.* **20**, 448-455; T. T. Yang, *et al.*, (1996) *Gene* **173**, 19-23), or an enzyme such as alkaline phosphatase which has been previously displayed on gIII (J. McCafferty, R. H. Jackson, D. J. Chiswell, (1991) *Protein Engineering* **4**, 955-961) Labelling different types of diagnostic phage with distinct markers would allow multiplex screening of a single DNA sample. Nevertheless, even in the absence of such refinements, the basic ELISA technique is reliable, fast, simple and particularly inexpensive. Moreover it requires no specialised apparatus, nor does it employ hazardous reagents such as radioactive isotopes, making it amenable to routine use in the clinic. The major advantage of the protocol is that it obviates the requirement for gel electrophoresis, and so opens the way to automated DNA diagnosis.

The invention provides DNA binding proteins which can be engineered with high specificity. The invention lends itself, therefore, to the design of any molecule of which specific DNA binding is required. For example, the proteins according to the invention may be employed in the manufacture of chimeric restriction enzymes, in which a nucleic acid cleaving domain is fused to a DNA binding domain comprising a zinc finger as described herein.

Preferably, the DNA binding molecules of the invention may bind the target nucleic acid with different affinity in the prescence or in the abscence of ligand. The binding

to the nucleic acid may be enhanced by the prescence of the ligand (ie bind with a higher affinity in the prescence of ligand), or may be reduced in the prescence of ligand (ie bind with a lower affinity in the prescence of ligand). In the case where association of the DNA binding molecule(s) with the target nucleic acid is enhanced by the prescence of ligand, said association may be additive with the binding of the ligand, or may be synergistic with the binding of the ligand, or may affect the binding in another way. If the binding is synergistic with the binding of the ligand, said binding may be either wholly or partly dependent on the prescence of the ligand. Preferably, the characteristics of binding may be such that the DNA binding molecule(s) may be eluted by addition of a molar excess of the DNA binding ligand.

Molecules according to the invention are preferably polypeptide sequences, optionally encoded by nucleic acid sequences. Fragments, mutants, alleles and other derivatives of the molecules of the invention preferably retain substantial homology with said sequence(s). As used herein, "homology" means that the two entities share sufficient characteristics for the skilled person to determine that they are similar. Preferably, homology is used to refer to sequence identity. Thus, the derivatives of said DNA binding molecules of the invention preferably retain substantial sequence identity with said molecules.

In the context of the present invention, a homologous sequence is taken to include any sequence which is at least 60, 70, 80 or 90% identical, preferably at least 95 or 98% identical over at least 5, preferably 8, 10, 15, 20, 30, 40 or even more residues or bases with the molecules (ie. the sequences thereof) of the invention, for example as shown in the sequence listing herein. In particular, homology should typically be considered with respect to those regions of the molecule(s) which may be known to be functionally important rather than non-essential neighbouring sequences. Homology comparisons can be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. In some aspects of the present invention, no gap penalties are used when determining sequence identity.

Relative sequence identity may be determined by computer programs which can calculate the percentage identity between two or more sequences using any suitable algorithm for determining identity, using for example default parameters. A typical example of such a computer program is CLUSTAL (see Thompson *et al.*, 1994 (NAR 22:4673-80) or http://www.psc.edu/general/software/packages/clustal/clustal.html). Advantageously, the BLAST algorithm is employed, with parameters set to default values. The BLAST algorithm is described in detail at http://www.ncbi.nih.gov/BLAST/blast_help.html, which is incorporated herein by reference.

Other computer programs used to determine identity and/or similarity between sequences include but are not limited to the GCG program package (Devereux *et al* 1984 Nucleic Acids Research 12:387), FASTA (Atschul *et al* 1990 J Mol Biol 403-410) and the GENEWORKS suite of comparison tools.

FASTA uses the method of Pearson and Lipman (Proc. Natl. Acad. Sci. USA 85; 2444-2448 (1988)) to search for similarities between one sequence (the query) and any group of sequences. FASTA uses the following search parameters: these can be advantageously set to the defined default parameters: Matrix: as for BLAST (not used by FASTA for nucleotide comparisons). Wordsize - the number of continuous residues or bases which are considered at once in the initial comparison; default is 6 for nucleotide sequences, 2 for amino acid sequences. Gap penalty: This is the number of points deducted from a similarity score when a new gap is created; default is 16 for nucleotide sequences, 12 for amino acid sequences. Gap extension penalty: This is the number of points deducted from a similarity score when an existing gap is enlarged; default is 4 for nucleotide sequences, 2 for amino acid sequences. Expect: this restricts the number of sequences returned according to statistical significance; default is 2. List: this restricts the number of homologous sequences which are reported; default is 40. Align: this restricts the number of homologous sequences for which alignments are displayed; default is 10.

FASTA is freely available *via* Biology WorkBench at the University of Illinois (http://biology.ncsa.uiuc.edu/), or through the SEQNET facility at (http://www.seqnet.dl.ac.uk//dbsearch.html).

BLAST (Basic Local Alignment Search Tool) is a heuristic search algorithm employed by the programs blastp, blastn, blastx, tblastn, and tblastx; these programs ascribe significance to their findings using the statistical methods of Karlin and Altschul (see http://www.ncbi.nih.gov/BLAST/blast_help.html) with a few enhancements. The BLAST programs were tailored for sequence similarity searching, for example to identify homologues to a query sequence. For a discussion of basic issues in similarity searching of sequence databases, see Altschul *et al* (1994:Nature Genetics 6:119-29).

BLAST uses the following search parameters: these can be advantageously set to the defined default parameters: HISTOGRAM - Displays a histogram of scores for each search; default is yes. DESCRIPTIONS - Restricts the number of descriptions of homologous sequences reported; default is 100. EXPECT - The statistical significance threshold for matches between sequences, according to the stochastic model of Karlin and Altschul (1990: PNAS 87:2264-8); default is 10. ALIGNMENTS - Restricts the number of sequences for which alignments are displayed; default is 50. MATRIX - Specifies a scoring matrix for BLASTP, BLASTX, TBLASTN and TBLASTX. The default matrix is BLOSUM62 (Henikoff & Henikoff 1992:PNAS 89:10915-9). STRAND - Restrict a search to one or other strands of the sequence, (if a nucleotide sequence); default is both strands. FILTER - Masks off segments of the query sequence which have low complexity, as determined by the SEG program of Wootton & Federhen (1993: Computers in Chemistry 17:149-163), or segments consisting of short-periodicity internal repeats, as determined by the XNU program of Claverie & States (1993: Computers and Chemistry 17:191-201) or by the DUST program of Tatusov and Lipman (see http://www.ncbi.nlm.nih.gov); default filtering is DUST for BLASTN, SEG for other programs.

Most preferably, sequence comparisons are conducted using the simple BLAST search algorithm. This is readily available, for example at http://www.ncbi.nlm.nih.gov/BLAST.

Molecules according to the invention may include any atom, ion, molecule, macromolecule (for example polypeptide), or combination of such entities. Advantageously, molecules according to the invention may include families of polypeptides with known or suspected nucleic acid binding motifs. These may include for example zinc finger proteins (see above). Molecules according to the invention may also include helix-turn-helix proteins, homeodomains, leucine zipper proteins, helix-loop-helix proteins or β-sheet motifs which are well known to a person skilled in the art.

According to the invention, DNA-binding motifs of one or more known or suspected nucleic acid binding polypeptide(s) may advantageously be randomised, in order to provide libraries of candidate nucleic acid binding molecules.

Crystal structures may advantageously be used in selecting or predicting the relevant DNA-binding regions of nucleic acid binding proteins by methods known in the art.

DNA-binding regions of proteins within the same structural family are often conserved or homologous to one another, for example zinc finger α-helices, the leucine zipper basic region, homeodomain helix 3.

General considerations and rules governing the binding of several polypeptide families to nucleic acids are set out in the literature, e.g. in (Suzuki *et* al., 1994:PNAS vol 91 pp 12357-61). Nucleic acid binding criteria for zinc fingers as preferred DNA binding molecules according to the present invention are set out in this application (see above).

It is also envisaged that the methods of the present invention could be advantageously applied to the selection of ligand-modulatable DNA binding molecules from other

families of transcription factors, for example from the helix-turn-helix (HTH) family and/or from the probe helix (PH) family, and/or from the C4 Zinc-binding family (which includes the hormone receptor (HR) family), from the Gal4 family, from the c-myb family, from other zinc finger families, or from any other family of DNA binding proteins known to one skilled in the art.

One or more polypeptides from one or more of these families could be advantageously randomised to provide a library of candidate molecules for use in the methods of the invention. Preferably, the amino acid residues known to be important for nuclei acid binding could be randomised.

The recognition helix of PH family polypeptides contains conserved Arg/Lys residues which are important structural elements involved in the binding of phosphates in the nucleic acid. Base specificity is attributed to amino acids 1,4,5 and 8 of the helix. These residues could be advantageously varied, for example amino acid 1 could be selected from Asn, Asp, His, Val, Ile to provide the possibility of binding to A,C,G, or T. Similarly, amino acid 4 could be selected from Asn, Asp, His, Val, Ile, Gln, Glu, Arg, Lys, Met, or Leu to provide the possibility of binding to A,C,G or T. Preferably, the rules laid out in (Suzuki *et* al., 1994:PNAS vol 91 pp 12357-61) would be used in order to randomise those amino acids which affect interaction of the molecule with the nucleic acid, whether in a base specific manner, or via binding to the phosphate backbone, thereby producing a library of candidate nucleic acid binding molecules for use in the methods of the invention.

Similarly, polypeptide molecules of the helix-turn-helix family could be randomised to produce a library of candidate molecules, at least some of which may preferably be capable of binding nucleic acid in a ligand-dependent manner when used in the methods of the present invention. In particular, amino acids 1,2,5 and 6 are known to be conserved and function in base-specific nucleic acid binding in HTH motifs. Therefore, at least amino acids 1,2,5 or 6 would preferably be randomised so as to produce molecules for use according to the present invention. More preferably, amino

acids 1,5 and 6 could be selected from Asn, Asp, His, Val, Ile, Glu, Gln, Arg, Met, Lys or Leu, and amino acid 2 could be selected from from Asn, Asp, His, Val, Ile, Glu, Gln, Arg, Met, Lys, Leu, Cys, Ser, Thr, or Ala.

Another family of transcription factors which may be advantageously employed in the methods of the current invention are the C4 family which includes hormone receptor type transcription factors. It is envisaged that polypeptides of this family could advantageously be used to provide candidate molecules for use in selecting nucleic acid binding molecules whose association with nucleic acid is modulatable by a nucleic acid binding ligand. Amino acids 1,4,5 and 9 of the C4 motif are known to be involved in contacting the DNA, and therefore these residues would preferably be altered to provide a plurality of different molecules which may bind DNA in a ligand dependent manner. Preferably, amino acids 1 and 5 could be selected from from Asn, Asp, His, Val, Ile, Glu, Gln, Arg, Met, Lys or Leu, and amino acids 4 and 9 could be selected from Gln, Glu, Arg, Lys, Leu or Met.

It is envisaged that the methods of the invention may be applied *in vivo*, for example they could be applied to the selection or isolation of DNA binding molecules capable of associating with target DNA *in vivo* inside one or more cells, in a manner analagous to the one-hybrid system.

It is envisaged that the methods of the invention may be practised in parallel. For example, multiple target DNAs could be used in a single selective step, thereby enabling multiple DNA binding molecules to be isolated simultaneously, even in the same physical vessel. Said multiple DNA binding molecules may preferably be different from one another. Said multiple DNA binding molecules may have similar or identical DNA binding specificities, or may preferably have different DNA binding specificities.

The invention may be worked using multiple DNA binding ligands, either seperately or in combination. For example, a target nucleic acid sequence may be used to isolate

DNA binding molecules according to the methods essentially as disclosed above, with the modification that more than one DNA binding ligand may be present. In this way, it is possible to isolate multiple DNA binding molecules which require different ligands to bind to the same target nucleic acid sequence(s).

It is further envisaged that the methods of the invention may be advantageously used to select nucleic acid sequences which allow binding of a particular DNA binding ligand/DNA binding molecule combination. For example, one may wish to isolate particular DNA sequences to which a given DNA binding molecule is able to bind, or to isolate only those DNA sequences which depend on the prescence of ligand for the DNA binding molecule to associate with them.

Accordingly, there is provided a method for isolating target DNA sequences to which a particular DNA binding molecule will bind, said method comprising

a) providing a library of target nucleic acid molecule(s);
b) contacting said nucleic acid molecules with a DNA binding molecule in the prescence or abscence of DNA binding ligand
c) assessing the ability of the candidate target DNA molecule(s) to bind the DNA binding molecule; and
d) isolating those target nucleic acid molecules which bind the DNA binding molecule.

A library of target nucleic acid molecule(s) according to the invention may preferably comprise a plurality of different nucleic acid molecules; preferably said nucleic acid molecules may be related to one another in terms of sequence homology.

A library of candidate nucleic acid binding molecule(s) according to the invention may preferably comprise a plurality of different candidate nucleic acid binding polypeptides; preferably said candidate nucleic acid binding polypeptides may be related to one another in terms of amino acid sequence homology.

It is envisaged that this method could be advantageously used in order to isolate DNA sequences which require ligand to associate with a known DNA binding molecule. For example, there may be a DNA sequence which is bound by a known DNA binding molecule in a ligand-independent manner, and it may be desirable to find a DNA sequence(s) which can also associate with the same wild-type DNA binding molecule, but which do so in a ligand-modulatable manner. Preferably, this may be accomplished according to the above method of the present invention.

In a preferred embodiment, the invention provides a method for isolating multiple DNA binding molecules in the prescence of multiple DNA binding ligands, said DNA binding molecules being selected using one or more target nucleic acid sequences in a single selection (isolation) procedure.

Accordingly, a method is provided for isolating one or more nucleic acid binding molecules, said molecules each binding one or more target nucleic acid sequence(s), wherein said binding to one or more target nucleic acid sequence(s) is modulatable by one or more nucleic acid binding ligands, and wherein said nucleic acid binding ligands and said nucleic acid binding molecule(s) are different, said method comprising:

a) providing one or more target nucleic acid molecule(s);

b) contacting the target nucleic acid molecule(s) with one or more nucleic acid binding ligand(s), to produce one or more nucleic acid-ligand complex(es);

c) assessing the ability of candidate nucleic acid binding molecules to bind the target nucleic acid molecule(s) and the nucleic acid-ligand complex(es); and

d) isolating those candidate nucleic acid binding molecules which bind the target nucleic acid molecule(s) and DNA-ligand complex(es) with different binding affinities.

In addition, the invention provides a method for isolating one or more DNA binding molecules, said molecules each binding one or more target DNA sequence(s), wherein said binding to one or more target DNA sequence(s) is modulatable by one or more DNA-binding ligands, and wherein said DNA-binding ligands and said DNA binding molecule(s) are different, said method comprising:

a) providing one or more target DNA molecule(s);
b) contacting the target DNA molecule(s) with one or more DNA binding ligand(s), to produce one or more DNA-ligand complex(es);
c) providing a library of candidate DNA-binding molecules,
d) assessing the ability of candidate DNA binding molecules to bind the target DNA molecule(s) and the DNA-ligand complex(es); and
e) isolating those candidate DNA binding molecules which bind the target DNA molecule(s) and DNA-ligand complex(es) with different binding affinities.

The library of candidate DNA-binding molecules is preferably a phage display library, individual candidate molecules of the library optionally being structurally related to zinc finger transcription factors (for example see Choo and Klug, (1994) PNAS (USA) 91:11163-67, which describes aspects of such libraries and is incorporated herein by reference). This library is preferably constructed with DNA sequences of the form GCGNNNGCG (where all 64 middle triplets are represented in the mixture).

One or more DNA binding ligands means at least one DNA binding ligand, preferably two, three or four DNA binding ligands, more preferably five, six, or seven DNA binding ligands, most preferably a mixture of eight DNA-binding ligands, or even more. The ligands may be in any molar ratio to one another within the mixture, but will preferably be approximately equimolar with one another.

The method would preferably feature a pre-selection step as described above to remove candidate DNA binding molecules which do not require ligand to bind the DNA.

Said method would preferably be carried out over about 6 rounds of selection.

DNA binding molecules (ie. phage clones) isolated by this method would preferably be individually assayed (for example in microtitre plates as described below) for binding to the mixture GCGNNNGCG in the presence and absence of a mixture of the DNA-binding ligands to identify clones which are capable of ligand-modulatable binding.

Those phage clones which are capable of ligand-modulatable binding would preferably be tested in the presence of a mixture of the eight ligands, in order to deduce the optimum target DNA sequence, for example using different or variant target DNA sequences, or by the binding site signature method method (see Choo and Klug, (1994) PNAS (USA) 91:11163-67).

In a further aspect of the invention, DNA binding molecules according to the invention may be advantageously used to determine the sequence composition of a sample of target DNA. For example, a DNA binding molecule according to the invention may be prepared which binds to a known target DNA sequence. By applying this molecule to, or contacting it with, one or more test DNA samples and monitoring its binding thereto, it is possible to determine whether said DNA sample(s) contain the cognate DNA recognition site of the DNA binding molecule, and therefore derive information about the nucleotide composition of said DNA test sample(s). Such analyses may be advantageously conducted using the binding site signature method (see Choo and Klug, (1994) PNAS (USA) 91:11163-67).

Individual phage clones could advantageously be assayed for binding of their cognate DNA sequence(s) in the presence or abscence of individual ligands, to monitor which particular ligand modulates binding.

Clearly, it may be that more than one ligand modulates binding of DNA binding molecules to their cognate DNA sequence(s). Preferably, individual DNA binding

molecules (ie. phage clones) may be assayed for binding to target DNA sequence(s) in the presence of discrete ligand mixtures, wherein each ligand mixture preferably contains a unique mixture of ligands. In this way, the particular ligands which may modulate binding of a particular DNA binding molecule to its cognate target DNA sequence may advantageously be determined. For example, if it is found that two mixtures - one lacking ligand X and the other lacking ligand Y - are incapable of inducing binding, then a mixture of ligands X and Y may have the effect of moduating the binding. This could advantageously be further investigated according to the methods of the invention as described herein.

It is envisaged that this invention may be advantageosly used in the isolation of a DNA binding ligand which is capable of modulating the association of a particular DNA binding molecule with its target DNA sequence. Accordingly, the invention provides a method for isolating one or more DNA binding ligands, said ligands each binding one or more target DNA sequence(s), wherein said binding to one or more target DNA sequence(s) modulates the binding of one or more DNA-binding molecules, and wherein said DNA-binding molecule(s) and said DNA binding ligands are different, said method comprising:

a) providing one or more target DNA molecule(s);
b) contacting the target DNA molecule(s) with one or more DNA binding molecule(s)
c) providing a library of candidate DNA-binding ligands,
d) assessing the ability of candidate DNA binding ligands to modulate the association of the DNA binding molecule(s) with the target DNA molecule(s); and
e) isolating those candidate DNA binding ligands which modulate the association of the DNA binding molecule(s) with the target DNA molecule(s).

It is envisaged that the methods of the current invention may be advantageously applied to the selection of molecules capable of binding nucleic acids other than DNA, for example RNA. Structural considerations of RNA binding molecules are discussed in Afshar *et al* (Afshar *et al*, 1999: Curr. Op. Biotech. vol 10 pages 59-63). In particular, ligands suitable for use in the methods of the invention as applied to RNA include those ligands described above, or may be selected from aminoglycosides and their derivatives such as paromomycin, neomycin (for examples see Park *et al.*, 1996: J. Am. Chem. Soc. vol 118 pp10150-10155); aminoglycoside mimetics (Tok and Rando 1998: J. Am. Soc. Chem. vol 120 pp 8279-8280); acridine derivatives (for examples see Hamy et al, 1998: Biochemistry vol 37 pp5086-5095); small peptides ('aptamers'); polycationic compounds (for examples see Wang *et al*, 1998: Tetrahedron 54 pp7955-7976) or any other nucleic acid binding molecules known to those skilled in the art. In a preferred embodiment, derivatives or libraries of said nucleic acid binding ligands may be prepared.

Accordingly, the present invention provides a method for isolating a RNA binding molecule which binds to a target RNA molecule in a manner modulatable by a RNA-binding ligand, wherein said RNA-binding ligand and said RNA-binding molecule are different, said method comprising; providing a target RNA molecule; contacting the target RNA molecule with a RNA-binding ligand, to produce a RNA-ligand complex; assessing the ability of candidate RNA-binding molecules to bind the target RNA molecule and the RNA-ligand complex; and isolating those candidate RNA-binding molecules which bind the target RNA molecule and RNA-ligand complex with different binding affinities.

The present invention will now be described by way of example, in which reference is made to:

Figure 1 which shows a graph

Figure 2 which shows a graph

Brief Description of the Figures

In slightly more detail,

Figure 1 shows a graph of the effect of drug (ligand) concentration on binding of two independent phage to specific DNA sequences.

Figure 2 shows a graph of phage binding (absorbance or O.D./optical density) vs DNA concentration for drug sensitive isolates in the absence of ligands.

The invention is described below, for the purpose of illustration only, in the following examples.

Examples

Example 1

Preparation and Screening of a Zinc Finger Phage Display Library

Selection Of Zinc Finger Phage Binding DNA Targets In The Prescence Of DNA Binding Ligand Distamycin A.

A powerful method of selecting DNA binding proteins is the cloning of peptides (Smith (1985) Science 228, 1315-1317), or protein domains (McCafferty *et al.*, (1990) Nature 348:552-554; Bass *et al.*, (1990) Proteins 8:309-314), as fusions to the minor coat protein (pIII) of bacteriophage fd, which leads to their expression on the tip of the capsid. A phage display library is created comprising variants of the middle finger from the DNA binding domain of Zif268.

Materials And Methods

Construction And Cloning Of Genes. In general, procedures and materials are in accordance with guidance given in Sambrook *et al.*, Molecular Cloning. A Laboratory Manual, Cold Spring Harbor, 1989. The gene for the Zif268 fingers (residues 333-420) is assembled from 8 overlapping synthetic oligonucleotides (see Choo and

Klug, (1994) PNAS (USA) 91:11163-67), giving *Sfi*I and *Not*I overhangs. The genes for fingers of the phage library are synthesised from 4 oligonucleotides by directional end to end ligation using 3 short complementary linkers, and amplified by PCR from the single strand using forward and backward primers which contain sites for *Not*I and *Sfi*I respectively. Backward PCR primers in addition introduce Met-Ala-Glu as the first three amino acids of the zinc finger peptides, and these are followed by the residues of the wild type or library fingers as required. Cloning overhangs are produced by digestion with *Sfi*I and *Not*I where necessary. Fragments are ligated to 1µg similarly prepared Fd-Tet-SN vector. This is a derivative of fd-tet-DOG1 (Hoogenboom *et al.*, (1991) Nucleic Acids Res. 19, 4133-4137) in which a section of the pelB leader and a restriction site for the enzyme *Sfi*I (underlined) have been added by site-directed mutagenesis using the oligonucleotide:

5' CTCCTGCAGTTGGACCTGTGCCATGGCCGGCTGGGCCGCATAGAATGG AACAACTAAAGC 3' (Seq ID No. 1)

which anneals in the region of the polylinker. Electrocompetent DH5α cells are transformed with recombinant vector in 200ng aliquots, grown for 1 hour in 2xTY medium with 1% glucose, and plated on TYE containing 15µg/ml tetracycline and 1% glucose.

The zinc finger phage display library of the present invention contains amino acid randomisations in putative base-contacting positions from the second and third zinc fingers of the three-finger DNA-binding domain of Zif268, and contains members that bind DNA of the sequence XXXXXGGCG where X is any base. Further details of the library used may be found in WO 98/53057 which is incorporated herein by reference. The DNA sequences AAAAAAGGCG and AAAAAAGGCGAAAAAA are used as selection targets in this example because short runs of adenines can cause intrinsic DNA bending - moreover, the structure of the bend can be disrupted by binding of the antibiotic distamycin A.

Phage Selection.

Bacterial colonies containing zinc finger phage libraries are transferred from plates to 200ml 2xTY medium (16g/litre Bactotryptone, 10g/litre Bactoyeast extract, 5g/litre NaCl) containing 50μM $ZnCl_2$ and 15 μg/ml tetracycline. Bacterial cultures are grown overnight at 30°C. Culture supernatant containing phages is obtained by centrifuging at 300xg for 5 minutes.

Phage selection is over 4 rounds. Before each round, a pre-selection step is included comprising binding of 10 pmol of biotinylated DNA target sites immobilised on 50mg streptavidin coated beads (Dynal) to 1 ml of phage solution (bacterial culture supernatant diluted 1:1 with PBS containing 50μM $ZnCl_2$, 4% Marvel, 2% Tween), for 1 hour at 20°C on a rolling platform. After this time, 0.5 ml of phage solution is transferred to a streptavidin coated tube and incubated with 2 pmol biotinylated DNA target site in the presence of 2μM distamycin A (Sigma) and 4μg poly [d(I-C)]. After a one hour incubation the tubes are washed 20 times with PBS containing 50μM $ZnCl_2$ and 1% Tween, and 3 times with PBS containing 50μM $ZnCl_2$. Phage are eluted using 0.1ml 0.1M triethylamine and the solution is neutralised with an equal volume of 1M Tris-Cl (pH 7.4). Logarithmic-phase *E. coli* TG1 cells are infected with eluted phage, and grown overnight, as described above, to prepare phage supernatants for subsequent rounds of selection.

After 4 rounds of selection, bacteria are plated and phage prepared from 96 colonies are screened for binding to the DNA target site in the presence and absence of distamycin A. Binding reactions are carried out in wells of a streptavidin-coated microtitre plate (Boehringer Mannheim) and contain 50μl of phage solution (bacterial culture supernatant diluted 1:1 with PBS containing 50μM $ZnCl_2$, 4% Marvel, 2% Tween), 0.15 pmol DNA target site and 0.25 μg poly [d(I-C)]. When added, distamycin A is present at a concentration of 2μM. After a one hour incubation the wells are washed 20 times with PBS containing 50μM $ZnCl_2$ and 1% Tween (and also distamycin A at a concentration of 2μM where appropriate), and 3 times with PBS containing 50μM $ZnCl_2$. Bound phage are detected by ELISA (carried out in the presence of distamycin A at a concentration of 2μM where appropriate) with

horseradish peroxidase-conjugated anti-M13 IgG (Pharmacia Biotech) and quantitated using SOFTMAX 2.32 (Molecular Devices).

Sequencing Of Selected Phage. Single colonies of transformants obtained after four rounds of selection as described, are grown overnight in 2xTY/Zn/Tet. Small aliquots of the cultures are stored in 15% glycerol at -20°C, to be used as an archive. Single-stranded DNA is prepared from phage in the culture supernatant and sequenced using the Sequenase™ 2.0 kit (U.S. Biochemical Corp.). The amino acid sequences of the zinc finger clones are deduced.

Amino acid sequences from helical regions of zinc fingers selected to bind DNA in the presence of distamycin

	F1	F2	F3
	-1123456	-1123456	-1123456
Clone 1	RSDELTR	RSDDLST	TNNTRIK
Clone 2	RSDELTR	RSDDLST	HKATRIK
Clone 3	RSDELTR	RSDDLST	TDKVRKK
Clone 4	RSDELTR	RSDDLST	HNASRIN
Clone 5	RSDELTR	RSDDLSV	TNNSRKK
Clone 6	RSDELTR	RSDDLST	TNATRKK
Clone 7	RSDELTR	RSDDLSQ	TRNTRKN
Clone 8	RSDELTR	RSDDLSV	TNNSRKN

Clones 1-4 were selected to bind the oligo:
tataAAAAAAGGCGTGtcacagtcagtccacacgtc

Clones 5-8 were selected to bind the oligo:
tataAAAAAAGGCGAAAAAAtcacagtcagtccacacgtc

Zinc finger phage clones are isolated according to this method which bind the DNA target with higher affinity in the presence of DNA binding ligand than in the abscence of DNA binding ligand (see Table 1, and Figures 1 and 2).

Table 1: Effect of ligand concentration on the binding of two independent phage clones to DNA sequences

[Ligand] (M)	actinomycin D isolate 1/30	distamycin A isolate 2/3			[target DNA]	actinomycin D isolate 1/30	Distamycin A isolate 2/3
0	0.49	0.811			0	0.122	0.131
0.000000001	0.562	0.825			0.15625	0.122	0.163
0.00000001	0.458	0.934			0.3125	0.164	0.237
0.0000001	0.43	0.771			0.625	0.187	0.281
0.000001	0.434	0.761			1.25	0.212	0.458
0.00001	0.269	0.751			2.5	0.424	0.613
0.00001	0.139	0.134			5	0.899	0.838
					10	1.202	1.101

Example 2

Modulation Of Binding Of Polypeptides To Target DNA By DNA Binding Ligand

Individual phage clones are assayed for modulation of target DNA binding by ligand in a phage ELISA binding assay.

Binding assay reactions are carried out in wells of a streptavidin-coated microtitre plate (Boehringer Mannheim) as in Example 1, except that the distamycin concentration is varied while the DNA concentration is kept constant at 2nM.

Induction of higher affinity DNA binding is observed when distamycin is added to the binding reaction at 10^{-6}M - 10^{-7}M.

Binding of the zinc finger phage to DNA in the absence of ligand, or at ligand concentrations of 10^{-9}M or lower, results in phage retention close to background level, ie. lower affinity binding than in the prescence of ligand.

Background level affinity binding is defined as the phage retention in binding reactions that contain no DNA binding site.

Example 3

DNA-Ligand Modulatable Restriction Enzyme

Phage-selected or rationally designed zinc finger domains which bind target DNA sequences in a manner modulatable by a DNA binding ligand can be converted to restriction enzymes which cleave DNA containing said target sequences in a manner modulatable by DNA binding ligand. This is achieved by coupling an appropriate zinc finger, as isolated in Example 1 above, to a cleavage domain of a restriction enzyme or other nucleic acid cleaving moiety.

A method of converting zinc finger DNA-binding domains to chimaeric restriction endonucleases has been described in Kim, *et al.*, (1996) Proc. Natl. Acad. Sci. USA 93:1156-1160. In order to demonstrate the applicability of DNA ligand-modulatable zinc fingers to restriction enzymes, a fusion is made between the catalytic domain of Fok I as described by Kim *et al.* and a zinc finger of Example 1. Fusion of the zinc finger nucleic acid-binding domain to the catalytic domain of Fok I restriction enzyme results in a novel endonuclease which cleaves DNA adjacent to the DNA recognition sequence of the zinc finger (AAAAAAGGCG or AAAAAAGGCGAAAAAA).

The oligonucleotides AAAAAAGGCG and AAAAAAGGCGAAAAAA are synthesised and ligated to arbitrary DNA sequences. After incubation with the zinc finger restriction enzyme, the nucleic acids are analysed by gel electrophoresis. Bands indicating cleavage of the nucleic acid at a position corresponding to the location of the oligonucleotide(s) (AAAAAAGGCG / AAAAAAGGCGAAAAAA) are visible.

In a further experiment, the zinc finger is fused to an amino terminal copper/nickel binding motif. Under the correct redox conditions (Nagaoka, M., *et al.*, (1994) J. Am. Chem. Soc. 116:4085-4086), sequence-specific DNA cleavage is observed, only in the presence of DNA incorporating oligonucleotide AAAAAAGGCG or AAAAAAGGCGAAAAAA.

Example 4

Modulation Of Transcriptional Activity In Vivo

A reporter system is produced which produces a reporter signal conditionally depending on the binding of the zinc finger DNA binding molecule to its target DNA sequence. This binding, and hence transcription from the reporter system, is modulated by the DNA binding ligand Distamycin A.

A transient transfection system using zinc finger transcription factors is produced as described in Choo, Y., *et al.*, (1997) J. Mol. Biol 273:525-532. This system comprises an expression plasmid which produces a phage-selected zinc finger fused to the activation domain of HSV VP16, and a reporter plasmid which contains the recognition sequence of the zinc finger upstream of a CAT reporter gene.

Thus, a zinc finger which recognises the DNA sequence AAAAAAGGCG is selected by phage display as described in Example 1. By the method of the preceding examples, said zinc finger is used to construct transcription factors as described above.

A transient expression experiment is conducted, wherein the CAT reporter gene on the reporter plasmid is placed downstream of the sequence AAAAAAGGCG. The reporter plasmid is cotransfected with a plasmid vector expressing the zinc finger-HSV fusion under the control of a constitutive promoter. No activation of CAT gene expression is observed.

However, when the same experiment is conducted in the presence of Distamycin A, CAT expression is observed as a result of the binding of the zinc finger transcription factor to its recognition sequence AAAAAAGGCG.

Example 5

Isolation of cognate target nucleic acids

Using a known DNA binding molecule, target DNA sequences to which it can bind are isolated.

The 434 repressor is a gene regulatory protein of phage 434. It binds to a 14bp operator site (see Koudelka *et al.*, 1987 Nature vol 326 pp 886-888). This operator site

consists of five conserved bp (1-5), then four variable bp (6-9), then five more conserved bp (10-14) as shown below:

```
Site:    1       5   6 7 8 9 10      14
Base:    A C A A G/T X X X X A/T T T G T
wherein X is any base.
```

The conserved bases contact the 434 repressor protein. The four variable bases are thought not to contact the 434 repressor protein. However, the four bases which do not contact the 434 repressor protein may affect the affinity of binding of the repressor to the operator site.

The 434 repressor protein (ie. the DNA binding molecule) is contacted with a library of different target DNA sequences in the prescence and abscence of ligand:

The target DNA sequences are synthesized using an Applied Biosystems 380A DNA synthesizer and are purified by gel electrophoresis. The four variable bases ('X' as shown above) are randomised, producing a library of 256 different target DNA molecules, position 5 being T, and position 10 being A. At the 5' and 3' ends of this sequence are placed PCR primer sequences for amplification and recovery of the central target sequences.

Structure of target DNA sequence library:

```
5'                        1    6  9    14                    3'
GTCGGATCCTGTCTGAGGTGAGACAATXXXXATTGTGTCTTCCGACGTCGAATTCGCG
wherein X is any base, and the partially randomised 434
operator is underlined.
```

The 434 repressor protein is added to the library of target DNA sequences, in the prescence and abscence of 2μM distamycin A (Sigma) ligand in 200μl binding buffer (9mM Tris-HCl pH 8.0, 90mM KCl, 90μM $ZnSO_4$) and incubated for 30 min.

Nitrocellulose filters (BA 85, Schleicher and Schüll) are placed into a suction chamber (as in Thiesen *et al. (eds)*, Immunological Methods vol IV, Academic Press, Orlando) and prewet with 600ml Tris-HCl binding buffer. The protein-oligonucleotide mix is applied to the filter(s) with gentle suction, the filters are washed with 4ml Tris-HCl binding buffer. Oligonucelotides are eluted in 200µl binding buffer plus 1mM 1-10-o-phenanthroline.

Oligonucleotides are then amplified by PCR, using the following primers:

Primer A 5'-GTCGGATCCTGTCTGAGGTGAG-3'
Primer B 5'-CGCGAATTCGACGTCGGAAGAC-3'

using an amplification kit (Perkin Elmer Cetus) with the following cycling regime:
93°C 30 sec
45°C 120 sec
45°C to 67°C ramp 60 sec
67°C 180 sec
for 25 cycles. 1µl of eluted oligonucleotide material is used as template.

Optionally, the PCR amplified DNA product is then used in further rounds of incubation with the 434 repressor protein, nitrocellulose filter binding, oligonucleotide elution and PCR amplification.

PCR amplified DNA products are then sequenced using standard techniques.

Target DNA sequences are selected which bind the 434 repressor with higher affinity in the prescence of ligand than in the abscence of ligand. Furthermore, DNA sequences are selected which bind the 434 repressor in the abscence of ligand with a higher affinity than in the prescence of ligand.

Example 6

Isolation of ligands which affect the binding of a DNA binding molecule to its cognate DNA target

The 434 repressor protein of Example 5 is used in conjunction with a target operator DNA sequence to which it binds.

The operator sequence used is
5'- A C A A T A A A T A T T G T -3'

A library of DNA binding ligands is used in place of the 2μM distamycin A (Sigma) DNA binding ligand of Example 5.

Ligands are isolated which are capable of increasing the affinity of the 434 repressor for its cognate DNA target sequence. Ligands are also isolated which are capable of decreasing the affinity of the 434 repressor for its cognate DNA target sequence.

CLAIMS

1) A method for isolating a DNA binding molecule which binds to a target DNA molecule in a manner modulatable by a DNA-binding ligand, wherein said DNA-binding ligand and said DNA-binding molecule are different, said method comprising:

a) providing a target DNA molecule;

b) contacting the target DNA molecule with a DNA-binding ligand, to produce a DNA-ligand complex;

c) assessing the ability of candidate DNA-binding molecules to bind the target DNA molecule and the DNA-ligand complex; and

d) isolating those candidate DNA-binding molecules which bind the target DNA molecule and DNA-ligand complex with different binding affinities.

2) A method for isolating a DNA binding molecule which binds to a target DNA molecule in a manner modulatable by a DNA-binding ligand, wherein said DNA-binding ligand and said DNA-binding molecule are different, and wherein said DNA-binding molecule has a higher affinity for the target DNA in the prescence of ligand than in the abscence of ligand, said method comprising:

a) providing a target DNA molecule;

b) contacting the target DNA molecule with a DNA-binding ligand, to produce a DNA-ligand complex;

c) assessing the ability of candidate DNA-binding molecules to bind the target DNA molecule and the DNA-ligand complex; and

d) isolating those candidate DNA-binding molecules which bind the DNA-ligand complex with a higher affinity than they bind the target DNA molecule.

3) A method for isolating a DNA binding molecule which binds to a target DNA molecule in a manner modulatable by a DNA-binding ligand, wherein said DNA-binding ligand and said DNA-binding molecule are different, and wherein said DNA binding molecule binds the target DNA in the abscence of

ligand with a higher affinity than it binds the target DNA in the prescence of ligand, said method comprising:

a) providing a target DNA molecule;

b) contacting the target DNA molecule with a DNA-binding ligand, to produce a DNA-ligand complex;

c) assessing the ability of candidate DNA-binding molecules to bind the target DNA molecule and the DNA-ligand complex; and

d) isolating those candidate DNA-binding molecules which bind the target DNA molecule in the abscence of ligand with a higher affinity than they bind the DNA-ligand complex.

4) The method according to any preceeding claim, wherein said target DNA molecule comprises a library of nucleic acid sequences, said sequences being related to one another by sequence homology.

5) The method according to any preceeding claim, wherein said candidate molecules are polypeptides.

6) The method according to any preceeding claim, wherein said candidate molecules are polypeptides at least partly derived from transcription factors.

7) The method according to any preceeding claim, wherein said candidate molecules are derived from zinc finger transcription factors.

8) A method according to any preceeding claim, wherein the candidate molecules are selected from a phage display library.

9) A method according to any preceeding claim, wherein the DNA binding ligand is Distamycin A.

10) DNA binding molecules obtainable by the method of any preceeding claim.

11) A method of modulating the expression of one or more genes, said method comprising

a) isolating one or more DNA binding molecule(s) according to any previous claim, and

b) administering said DNA binding molecule(s) to a cell.

ABSTRACT

The invention relates to a method for isolating a DNA binding molecule which binds to a target DNA molecule in a manner modulatable by a DNA-binding ligand, wherein said DNA-binding ligand and said DNA-binding molecule are different, said method comprising; providing a target DNA molecule; contacting the target DNA molecule with a DNA-binding ligand, to produce a DNA-ligand complex; assessing the ability of candidate DNA-binding molecules to bind the target DNA molecule and the DNA-ligand complex; and isolating those candidate DNA-binding molecules which bind the target DNA molecule and DNA-ligand complex with different binding affinities.

Figure 1: EFFECT OF DRUG CONCENTRATION ON BINDING (O.D.) OF TWO PHAGE TO SPECIFIC DNA SEQUENCES

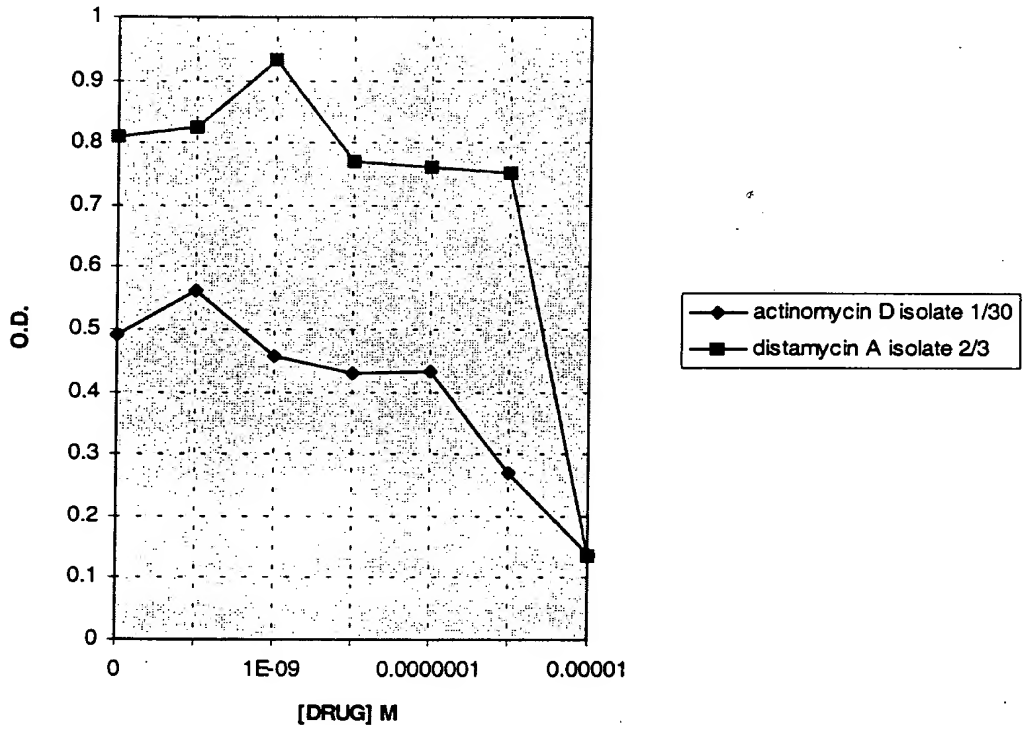

Figure 2: PHAGE BINDING (O.D.) VS. DNA CONCENTRATION FOR DRUG SENSITIVE ISOLATES IN THE ABSCENCE OF LIGAND(S)